The Theorem of Copeland and Erdős on Normal Numbers

Jordan Velich

University of Newcastle

February 3, 2015

Jordan Velich The Theorem of Copeland and Erdős on Normal Numbers

A number α is **normal** with respect to the base β , provided each of the digits $0, 1, 2, \ldots, \beta - 1$ occurs with a limiting relative frequency of $1/\beta$, and each of the β^k sequences of k digits occurs with the relative frequency $1/\beta^k$.

Definition

The **natural density** of the subset $A \subset \mathbb{N}$, denoted d(A), is defined as $N(A) = \frac{N(A)}{N(A)}$

$$d(A) = \lim_{x \to \infty} \frac{n(x)}{x}$$

where $N(x) := \#\{a : a \in A, a \leq x\}.$

Example

The natural density of the even numbers is 1/2.

A number α is **normal** with respect to the base β , provided each of the digits $0, 1, 2, \ldots, \beta - 1$ occurs with a limiting relative frequency of $1/\beta$, and each of the β^k sequences of k digits occurs with the relative frequency $1/\beta^k$.

Definition

The **natural density** of the subset $A \subset \mathbb{N}$, denoted d(A), is defined as N(x)

$$d(A) = \lim_{x \to \infty} \frac{n(x)}{x}$$

where $N(x) := \#\{a : a \in A, a \leq x\}.$

Example

The natural density of the even numbers is 1/2.

A number α is **normal** with respect to the base β , provided each of the digits $0, 1, 2, \ldots, \beta - 1$ occurs with a limiting relative frequency of $1/\beta$, and each of the β^k sequences of k digits occurs with the relative frequency $1/\beta^k$.

Definition

The **natural density** of the subset $A \subset \mathbb{N}$, denoted d(A), is defined as N(x)

$$d(A) = \lim_{x \to \infty} \frac{n(x)}{x}$$

where $N(x) := \#\{a : a \in A, a \leq x\}.$

Example

The natural density of the even numbers is 1/2.

Jordan Velich The Theorem of Copeland and Erdős on Normal Numbers

- Normality: The Known and Unkown
- Copeland-Erdős Theorem
- Questions of Strong Normality

Normality: The Known and Unkown

A number α is **normal** with respect to the base β , provided each of the digits $0, 1, 2, \ldots, \beta - 1$ occurs with a limiting relative frequency of $1/\beta$, and each of the β^k sequences of k digits occurs with the relative frequency $1/\beta^k$.

With this definition in mind, we begin with the unknown:

Conjecture (Borel, 1950)

All real irrational algebraic numbers are normal.

A number α is **normal** with respect to the base β , provided each of the digits $0, 1, 2, \ldots, \beta - 1$ occurs with a limiting relative frequency of $1/\beta$, and each of the β^k sequences of k digits occurs with the relative frequency $1/\beta^k$.

With this definition in mind, we begin with the unknown:

Conjecture (Borel, 1950)

All real irrational algebraic numbers are normal.

A number α is **normal** with respect to the base β , provided each of the digits $0, 1, 2, \ldots, \beta - 1$ occurs with a limiting relative frequency of $1/\beta$, and each of the β^k sequences of k digits occurs with the relative frequency $1/\beta^k$.

With this definition in mind, we begin with the unknown:

Conjecture (Borel, 1950)

All real irrational algebraic numbers are normal.

Some of the most **well-known normal numbers** discovered so far include:

1933	0.123456789
1946	0.23571113
1956	0.f(1)f(2)f(3)
1973	$\sum_{k=1}^{\infty} \frac{1}{b^{c^k} c^k}$

Champernowne's Number Copeland-Erdős Number Davenport-Erdős Numbers

Stoneham Numbers

Some of the most **well-known normal numbers** discovered so far include:

19330.123456789...19460.23571113...19560.f(1)f(2)f(3)...1973 $\sum_{k=1}^{\infty} \frac{1}{b^{c^k}c^k}$

Champernowne's Number Copeland-Erdős Number Davenport-Erdős Numbers Stoneham Numbers

Copeland-Erdős Theorem

Theorem (Copeland-Erdős, 1946)

If a_1, a_2, a_3, \ldots is an increasing sequence of integers such that for every $\theta < 1$ the number of a_i 's up to N exceeds N^{θ} provided N is sufficiently large, then the infinite decimal

0.*a*1*a*2*a*3...

is normal with respect to the base β in which these integers are expressed.

Lemma (Copeland-Erdős, 1946)

The number of integers up to N (N sufficiently large) which are not (ε, k) normal with respect to a given base β is less than N^{δ} where $\delta = \delta(\varepsilon, k, \beta) < 1$.

To understand this lemma, we must first be familiar with (ε, k) normality:

Definition

A number α (in the base β) is said to be (ε, k) normal if any combination of k digits appears consecutively among the digits of α with a relative frequency between $\beta^{-k} - \varepsilon$ and $\beta^{-k} + \varepsilon$.

Lemma (Copeland-Erdős, 1946)

The number of integers up to N (N sufficiently large) which are not (ε, k) normal with respect to a given base β is less than N^{δ} where $\delta = \delta(\varepsilon, k, \beta) < 1$.

To understand this lemma, we must first be familiar with (ε, k) normality:

Definition

A number α (in the base β) is said to be (ε, k) normal if any combination of k digits appears consecutively among the digits of α with a relative frequency between $\beta^{-k} - \varepsilon$ and $\beta^{-k} + \varepsilon$.

Lemma (Copeland-Erdős, 1946)

The number of integers up to N (N sufficiently large) which are not (ε, k) normal with respect to a given base β is less than N^{δ} where $\delta = \delta(\varepsilon, k, \beta) < 1$.

To understand this lemma, we must first be familiar with (ε, k) normality:

Definition

A number α (in the base β) is said to be (ε, k) normal if any combination of k digits appears consecutively among the digits of α with a relative frequency between $\beta^{-k} - \varepsilon$ and $\beta^{-k} + \varepsilon$.

Let x be such that $\beta^{x-1} \leq N < \beta^x$, where β^x refers to a number (base- β) consisting of x digits.

We introduce the notation $\beta_j = (\beta - 1)^{x-j} {X \choose j}$, where β_j counts the number of numbers (up to N) which have a single digit, for instance 0, occurring a total of j times amongst their x digits.

Since a given base β has β digits, and since β_j counts the occurrences of only a single digit, there are at most

$$\beta \left(\sum_{j < x(1-\varepsilon)/\beta} \beta_j + \sum_{j > x(1+\varepsilon)/\beta} \beta_j \right)$$

numbers up to N which have less than $x(1-\varepsilon)/\beta$ or more than $x(1+\varepsilon)/\beta$ 0's,1's,..., $(\beta-1)$'s.

Let x be such that $\beta^{x-1} \leq N < \beta^x$, where β^x refers to a number (base $-\beta$) consisting of x digits.

We introduce the notation $\beta_j = (\beta - 1)^{x-j} {\binom{x}{j}}$, where β_j counts the number of numbers (up to *N*) which have a single digit, for instance 0, occurring a total of *j* times amongst their *x* digits.

Since a given base β has β digits, and since β_j counts the occurrences of only a single digit, there are at most

$$\beta \left(\sum_{j < x(1-\varepsilon)/\beta} \beta_j + \sum_{j > x(1+\varepsilon)/\beta} \beta_j \right)$$

numbers up to N which have less than $x(1-\varepsilon)/\beta$ or more than $x(1+\varepsilon)/\beta$ 0's,1's,..., $(\beta-1)$'s.

Let x be such that $\beta^{x-1} \leq N < \beta^x$, where β^x refers to a number (base $-\beta$) consisting of x digits.

We introduce the notation $\beta_j = (\beta - 1)^{x-j} {\binom{x}{j}}$, where β_j counts the number of numbers (up to *N*) which have a single digit, for instance 0, occurring a total of *j* times amongst their *x* digits.

Since a given base β has β digits, and since β_j counts the occurrences of only a single digit, there are at most

$$\beta\left(\sum_{jx(1+\varepsilon)/\beta}\beta_j\right)$$

numbers up to N which have less than $x(1-\varepsilon)/\beta$ or more than $x(1+\varepsilon)/\beta$ 0's,1's,..., $(\beta-1)$'s.

In order to prove the lemma for $(\varepsilon, 1)$ normality, we have to show that

$$\beta\left(\sum_{jx(1+\varepsilon)/\beta}\beta_j\right)<\mathsf{N}^\delta.$$

We first require some intermediate inequalities.

We have from the properties of the binomial expansion:

$$\sum_{j < x(1-arepsilon)/eta} eta_j < (x+1)eta_{r_1}, \quad \sum_{j > x(1+arepsilon)/eta} eta_j < (x+1)eta_{r_2}$$

where

$$r_1 = \lfloor (1 - \varepsilon) x / \beta \rfloor, \quad r_2 = \lfloor (1 + \varepsilon) x / \beta \rfloor$$

In order to prove the lemma for $(\varepsilon, 1)$ normality, we have to show that

$$eta \left(\sum_{j < x(1-arepsilon)/eta} eta_j + \sum_{j > x(1+arepsilon)/eta} eta_j
ight) < N^{\delta}.$$

We first require some intermediate inequalities.

We have from the properties of the binomial expansion:

$$\sum_{j < x(1-arepsilon)/eta} eta_j < (x+1)eta_{r_1}, \quad \sum_{j > x(1+arepsilon)/eta} eta_j < (x+1)eta_{r_2}$$

where

$$r_1 = \lfloor (1-\varepsilon)x/\beta \rfloor, \quad r_2 = \lfloor (1+\varepsilon)x/\beta \rfloor$$

In order to prove the lemma for $(\varepsilon, 1)$ normality, we have to show that

$$\beta\left(\sum_{jx(1+\varepsilon)/\beta}\beta_j\right)<\mathsf{N}^\delta.$$

We first require some intermediate inequalities.

We have from the properties of the binomial expansion:

$$\sum_{j < \mathsf{x}(1-\varepsilon)/\beta} \beta_j < (\mathsf{x}+1)\beta_{r_1}, \quad \sum_{j > \mathsf{x}(1+\varepsilon)/\beta} \beta_j < (\mathsf{x}+1)\beta_{r_2}$$

where

$$r_1 = \lfloor (1 - \varepsilon) x / \beta \rfloor, \quad r_2 = \lfloor (1 + \varepsilon) x / \beta \rfloor$$

By repeated application of the relation

$$\beta_{j+1}/\beta_j = (x-j)/(j+1)(\beta-1)$$

we obtain

$$\beta_{r_1}\rho_1^{\varepsilon x/2} < \beta^x, \quad \beta_{r_2}\rho_2^{\varepsilon x/2} < \beta^x$$

where

$$\rho_1 = (x - r_1)/(r_1 + 1)(\beta - 1), \quad \rho_2 = (x - r_2)/(r_2 + 1)(\beta - 1)$$

and where $\rho_1, \rho_2 > 1$ for x sufficiently large.

It follows that

$$\beta_{r_1} < \left(\rho_1^{-\varepsilon/2}\beta\right)^{\times}, \quad \beta_{r_2} < \left(\rho_2^{-\varepsilon/2}\beta\right)^{\times}$$

By repeated application of the relation

$$\beta_{j+1}/\beta_j = (x-j)/(j+1)(\beta-1)$$

we obtain

$$\beta_{r_1}\rho_1^{\varepsilon x/2} < \beta^x, \quad \beta_{r_2}\rho_2^{\varepsilon x/2} < \beta^x$$

where

$$\rho_1 = (x - r_1)/(r_1 + 1)(\beta - 1), \quad \rho_2 = (x - r_2)/(r_2 + 1)(\beta - 1)$$

and where $\rho_1, \rho_2 > 1$ for x sufficiently large.

It follows that

$$\beta_{r_1} < \left(\rho_1^{-\varepsilon/2}\beta\right)^{\times}, \quad \beta_{r_2} < \left(\rho_2^{-\varepsilon/2}\beta\right)^{\times}$$

By repeated application of the relation

$$\beta_{j+1}/\beta_j = (x-j)/(j+1)(\beta-1)$$

we obtain

$$\beta_{r_1}\rho_1^{\varepsilon x/2} < \beta^x, \quad \beta_{r_2}\rho_2^{\varepsilon x/2} < \beta^x$$

where

$$\rho_1 = (x - r_1)/(r_1 + 1)(\beta - 1), \quad \rho_2 = (x - r_2)/(r_2 + 1)(\beta - 1)$$

and where $\rho_1, \rho_2 > 1$ for x sufficiently large.

It follows that

$$\beta_{r_1} < \left(\rho_1^{-\varepsilon/2}\beta\right)^x, \quad \beta_{r_2} < \left(\rho_2^{-\varepsilon/2}\beta\right)^x$$

Hence

$$\beta \left(\sum_{j < x(1-\varepsilon)/\beta} \beta_j + \sum_{j > x(1+\varepsilon)/\beta} \beta_j \right) < \beta(x+1) \left[\beta_{r_1} + \beta_{r_2} \right]$$
$$< \beta(x+1) \left[\left(\rho_1^{-\varepsilon/2} \beta \right)^x + \left(\rho_2^{-\varepsilon/2} \beta \right)^x \right]$$
$$< \beta^{\delta(x-1)} \leqslant N^{\delta}$$

and the lemma is established for (arepsilon,1) normality.

Hence

$$\beta \left(\sum_{j < x(1-\varepsilon)/\beta} \beta_j + \sum_{j > x(1+\varepsilon)/\beta} \beta_j \right) < \beta(x+1) \left[\beta_{r_1} + \beta_{r_2} \right]$$
$$< \beta(x+1) \left[\left(\rho_1^{-\varepsilon/2} \beta \right)^x + \left(\rho_2^{-\varepsilon/2} \beta \right)^x \right]$$
$$< \beta^{\delta(x-1)} \leqslant N^{\delta}$$

and the lemma is established for (arepsilon,1) normality.

Hence

$$\beta \left(\sum_{j < x(1-\varepsilon)/\beta} \beta_j + \sum_{j > x(1+\varepsilon)/\beta} \beta_j \right) < \beta(x+1) \left[\beta_{r_1} + \beta_{r_2} \right]$$
$$< \beta(x+1) \left[\left(\rho_1^{-\varepsilon/2} \beta \right)^x + \left(\rho_2^{-\varepsilon/2} \beta \right)^x \right]$$
$$< \beta^{\delta(x-1)} \leq N^{\delta}$$

and the lemma is established for (arepsilon,1) normality.

Hence

$$\beta \left(\sum_{j < x(1-\varepsilon)/\beta} \beta_j + \sum_{j > x(1+\varepsilon)/\beta} \beta_j \right) < \beta(x+1) \left[\beta_{r_1} + \beta_{r_2} \right]$$
$$< \beta(x+1) \left[\left(\rho_1^{-\varepsilon/2} \beta \right)^x + \left(\rho_2^{-\varepsilon/2} \beta \right)^x \right]$$
$$< \beta^{\delta(x-1)} \leqslant N^{\delta}$$

and the lemma is established for $(\varepsilon, 1)$ normality.

Consider the digits b_0, b_1, \ldots of a number $m \leq N$ grouped as follows:

 $b_0, b_1, \ldots, b_{k-1}; b_k, \ldots, b_{2k-1}; b_{2k}, \ldots, b_{3k-1}; \ldots$

Each of these groups represents a single digit of m when m is expressed in the base β^k . Hence, there are at most N^{δ} integers $m \leq N$ for which the frequency among these groups of a given combination of k digits falls outside the interval from $\beta^{-k} - \varepsilon$ to $\beta^{-k} + \varepsilon$.

The same holds for

and so on. This gives our result.

Consider the digits b_0, b_1, \ldots of a number $m \leq N$ grouped as follows:

$$b_0, b_1, \ldots, b_{k-1}; b_k, \ldots, b_{2k-1}; b_{2k}, \ldots, b_{3k-1}; \ldots$$

Each of these groups represents a single digit of m when m is expressed in the base β^k . Hence, there are at most N^{δ} integers $m \leq N$ for which the frequency among these groups of a given combination of k digits falls outside the interval from $\beta^{-k} - \varepsilon$ to $\beta^{-k} + \varepsilon$.

The same holds for

and so on. This gives our result.

Consider the digits b_0, b_1, \ldots of a number $m \leq N$ grouped as follows:

$$b_0, b_1, \ldots, b_{k-1}; b_k, \ldots, b_{2k-1}; b_{2k}, \ldots, b_{3k-1}; \ldots$$

Each of these groups represents a single digit of m when m is expressed in the base β^k . Hence, there are at most N^{δ} integers $m \leq N$ for which the frequency among these groups of a given combination of k digits falls outside the interval from $\beta^{-k} - \varepsilon$ to $\beta^{-k} + \varepsilon$.

The same holds for

$$b_1, b_2, \ldots, b_k; b_{k+1}, \ldots, b_{2k}; \ldots$$

and so on. This gives our result.

Theorem (Copeland-Erdős, 1946)

If a_1, a_2, a_3, \ldots is an increasing sequence of integers such that for every $\theta < 1$ the number of a_i 's up to N exceeds N^{θ} provided N is sufficiently large, then the infinite decimal

0.*a*1*a*2*a*3...

is normal with respect to the base β in which these integers are expressed.

Consider the numbers a_1, a_2, \ldots of the increasing sequence up to the largest $a_i \leq N$, where $N = \beta^n$. By a counting argument, these numbers altogether have at least $n(1 - \varepsilon) \cdot (N^{\theta} - N^{1-\varepsilon})$ digits.

Let f_N be the relative frequency of the digit 0. It follows from the lemma that the number of a_i 's for which the frequency of the digit 0 exceeds $\beta^{-1} + \varepsilon$ is at most N^{δ} , and hence

$$f_{N} < \beta^{-1} + \varepsilon + \frac{nN^{\delta}}{n(1-\varepsilon)(N^{\theta} - N^{1-\varepsilon})} \\ = \beta^{-1} + \varepsilon + \frac{N^{\delta-\theta}}{(1-\varepsilon)(1-N^{1-\varepsilon-\theta})}$$

Consider the numbers a_1, a_2, \ldots of the increasing sequence up to the largest $a_i \leq N$, where $N = \beta^n$. By a counting argument, these numbers altogether have at least $n(1 - \varepsilon) \cdot (N^{\theta} - N^{1-\varepsilon})$ digits.

Let f_N be the relative frequency of the digit 0. It follows from the lemma that the number of a_i 's for which the frequency of the digit 0 exceeds $\beta^{-1} + \varepsilon$ is at most N^{δ} , and hence

$$egin{split} f_{\mathcal{N}} &< eta^{-1} + arepsilon + rac{nN^{\delta}}{n(1-arepsilon)(N^{ heta}-N^{1-arepsilon})} \ &= eta^{-1} + arepsilon + rac{N^{\delta- heta}}{(1-arepsilon)(1-N^{1-arepsilon- heta})} \end{split}$$

Since we are permitted to take θ greater than δ and greater than $1 - \varepsilon$, it follows that $\lim_{N \to \infty} f_N$ is at most $\beta^{-1} + \varepsilon$ and hence at most β^{-1} .

A similar result holds for the digits $1, 2, ..., \beta - 1$ and hence each of these digits must have a limiting relative frequency of exactly β^{-1} .

In a similar manner, it can be shown that the limiting relative frequency of any combination of k digits is β^{-k} .

Since we are permitted to take θ greater than δ and greater than $1 - \varepsilon$, it follows that $\lim_{N \to \infty} f_N$ is at most $\beta^{-1} + \varepsilon$ and hence at most β^{-1} .

A similar result holds for the digits $1, 2, \ldots, \beta - 1$ and hence each of these digits must have a limiting relative frequency of exactly β^{-1} .

In a similar manner, it can be shown that the limiting relative frequency of any combination of k digits is β^{-k} .

Since we are permitted to take θ greater than δ and greater than $1 - \varepsilon$, it follows that $\lim_{N \to \infty} f_N$ is at most $\beta^{-1} + \varepsilon$ and hence at most β^{-1} .

A similar result holds for the digits $1, 2, \ldots, \beta - 1$ and hence each of these digits must have a limiting relative frequency of exactly β^{-1} .

In a similar manner, it can be shown that the limiting relative frequency of any combination of k digits is β^{-k} .

Questions of Strong Normality

A number α is **simply strongly normal** to the base β , if for each $k \in \{0, 1, \dots, \beta - 1\}$, we have

$$\limsup_{n \to \infty} \frac{m_k(n) - n/\beta}{\sqrt{2n \log \log n}} = \frac{\sqrt{\beta - 1}}{\beta} \quad \text{and}$$
$$\lim_{n \to \infty} \inf_{n \to \infty} \frac{m_k(n) - n/\beta}{\sqrt{2n \log \log n}} = -\frac{\sqrt{\beta - 1}}{\beta}$$

where $m_k(n) := \#\{i : a_i = k, i \leq n\}$. A number is **strongly normal** to the base β if it is simply strongly normal in each base $\beta^j, j = 1, 2, 3, ...$, and is **absolutely strongly normal** if it is strongly normal to every base.

A number α is **normal** with respect to the base β , if for each combination of k digits, $a_1a_2 \dots a_k$, we have

$$\lim_{x\to\infty}\frac{N(x)}{x}=\frac{1}{\beta^k}$$

where N(x) is the number of occurrences of $a_1a_2...a_k$ in the first x digits of α .

- A number which is strongly normal to the base β is normal to the base β.
- Champernowne's base- β number is not strongly normal to the base β .

A number α is **normal** with respect to the base β , if for each combination of k digits, $a_1a_2 \dots a_k$, we have

$$\lim_{x\to\infty}\frac{N(x)}{x}=\frac{1}{\beta^k}$$

where N(x) is the number of occurrences of $a_1a_2...a_k$ in the first x digits of α .

- A number which is strongly normal to the base β is normal to the base β.
- Champernowne's base- β number is not strongly normal to the base β .

A number α is **normal** with respect to the base β , if for each combination of k digits, $a_1a_2 \dots a_k$, we have

$$\lim_{x\to\infty}\frac{N(x)}{x}=\frac{1}{\beta^k}$$

where N(x) is the number of occurrences of $a_1a_2...a_k$ in the first x digits of α .

- A number which is strongly normal to the base β is normal to the base β.
- Champernowne's base- β number is not strongly normal to the base β .

A number α is **normal** with respect to the base β , if for each combination of k digits, $a_1a_2 \dots a_k$, we have

$$\lim_{x\to\infty}\frac{N(x)}{x}=\frac{1}{\beta^k}$$

where N(x) is the number of occurrences of $a_1a_2...a_k$ in the first x digits of α .

- A number which is strongly normal to the base β is normal to the base β.
- Champernowne's base- β number is not strongly normal to the base β .

Conjecture

The number α (= 0. $a_1a_2a_3...$) formed from the concatenation of the increasing sequence $a_1, a_2, a_3,...$ is not strongly normal, provided that the sequence of integers is dense enough, that is, $N(x) > x^{\theta}$ for every $\theta < 1$ and sufficiently large x.

• Heuristic: This conjecture is put forward as a consequence of Champernowne's number failing to be strongly normal. We believe that all the other concatenation numbers should also fail this strong normality test, the reason being that these sequences are just too dense – there are too few integers being excluded. In this way, we see these concatenation numbers as being basically the same as Champernowne's number – too structured – and thus we conjecture that these numbers should fail to be strongly normal.

Conjecture

The number α (= 0. $a_1a_2a_3...$) formed from the concatenation of the increasing sequence $a_1, a_2, a_3,...$ is not strongly normal, provided that the sequence of integers is dense enough, that is, $N(x) > x^{\theta}$ for every $\theta < 1$ and sufficiently large x.

• Heuristic: This conjecture is put forward as a consequence of Champernowne's number failing to be strongly normal. We believe that all the other concatenation numbers should also fail this strong normality test, the reason being that these sequences are just too dense – there are too few integers being excluded. In this way, we see these concatenation numbers as being basically the same as Champernowne's number – too structured – and thus we conjecture that these numbers should fail to be strongly normal.

Thank you!